How GPS Receivers Work

by Marshall Brain and Tom Harris
www.howstuffworks.com
26 September 2006

Our ancestors had to go to pretty extreme measures to keep from getting lost. They erected monumental landmarks, laboriously drafted detailed maps and learned to read the stars in the night sky.

Things are much, much easier today. For less than $100, you can get a pocket-sized gadget that will tell you exactly where you are on Earth at any moment. As long as you have a GPS receiver and a clear view of the sky, you'll never be lost again.
In this article, we'll find out how these handy guides pull off this amazing trick. As we'll see, the Global Positioning System is vast, expensive and involves a lot of technical ingenuity, but the fundamental concepts at work are quite simple and intuitive.
When people talk about "a GPS," they usually mean a GPS receiver. The Global Positioning System (GPS) is actually a constellation of 27 Earth-orbiting satellites (24 in operation and three extras in case one fails). The U.S. military developed and implemented this satellite network as a military navigation system, but soon opened it up to everybody else.

Each of these 3,000- to 4,000-pound solar-powered satellites circles the globe at about 12,000 miles (19,300 km), making two complete rotations every day. The orbits are arranged so that at any time, anywhere on Earth, there are at least four satellites "visible" in the sky.

A GPS receiver's job is to locate four or more of these satellites, figure out the distance to each, and use this information to deduce its own location. This operation is based on a simple mathematical principle called trilateration. Trilateration in three-dimensional space can be a little tricky, so we'll start with an explanation of simple two-dimensional trilateration.

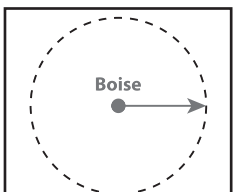


Figure 1

2-D Trilateration
Imagine you are somewhere in the United States and you are TOTALLY lost -- for whatever reason, you have absolutely no clue where you are. You find a friendly local and ask, "Where am I?" He says, "You are 625 miles from Boise, Idaho." This is a nice, hard fact, but it is not particularly useful by itself. You could be anywhere on a circle around Boise that has a radius of 625 miles (See Figure 1).

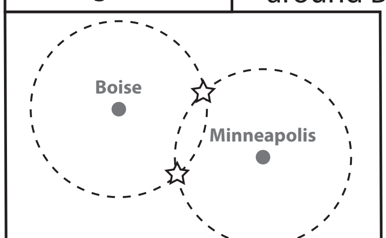


Figure 2

You ask somebody else where you are, and she says, "You are 690 miles from Minneapolis, Minnesota." Now you're getting somewhere. If you combine this information with the Boise information, you have two circles that intersect. You now know that you must be at one of these two intersection points, if you are 625 miles from Boise and 690 miles from Minneapolis (See Figure 2).

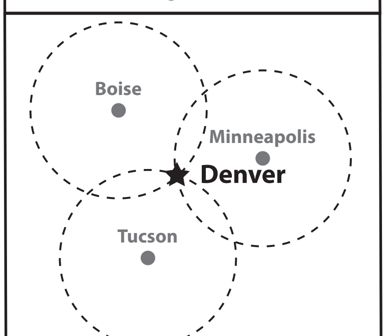
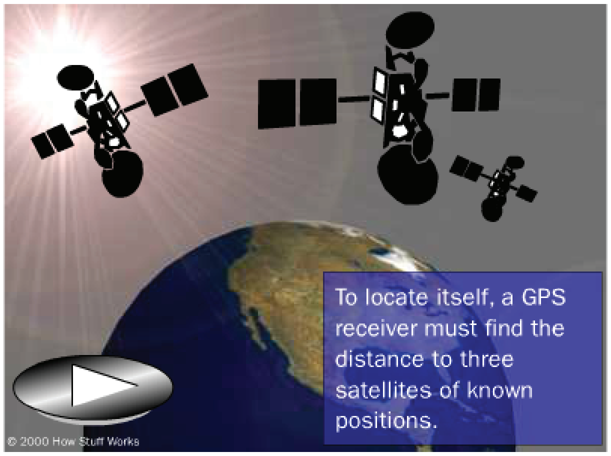


Figure 3

If a third person tells you that you are 615 miles from Tucson, Arizona, you can eliminate one of the possibilities, because the third circle will only intersect with one of these points. You now know exactly where you are -- Denver, Colorado.

This same concept works in three-dimensional space, as well, but you're dealing with spheres instead of circles. In the next section, we'll look at this type of trilateration.

3-D Trilateration



Fundamentally, three-dimensional trilateration isn't much different from two-dimensional trilateration, but it's a little trickier to visualize. Imagine the radii from the previous examples going off in all directions. So instead of a series of circles, you get a series of spheres. If you know you are 10 miles from satellite A in the sky, you could be anywhere on the surface of a huge, imaginary sphere with a 10-mile radius. If you also know you are 15 miles from satellite B, you can overlap the first sphere with another, larger sphere. The spheres intersect in a perfect circle. If you know the distance to a third satellite, you get a third sphere, which intersects with this circle at two points.

The Earth itself can act as a fourth sphere -- only one of the two possible points will actually be on the surface of the planet, so you can eliminate the one in space. Receivers generally look to four or more satellites, however, to improve accuracy and provide precise altitude information.

In order to make this simple calculation, then, the GPS receiver has to know two things:

- The location of at least three satellites above you
- The distance between you and each of those satellites

The GPS receiver figures both of these things out by analyzing high-frequency, low-power radio signals from the GPS satellites. Better units have multiple receivers, so they can pick up signals from several satellites simultaneously.

Radio waves are electromagnetic energy, which means they travel at the speed of light (about 186,000 miles per second, 300,000 km per second in a vacuum). The receiver can figure out how far the signal has traveled by timing how long it took the signal to arrive. In the next section, we'll see how the receiver and satellite work together to make this measurement.

GPS Calculations

In the previous section, we saw that a GPS receiver calculates the distance to GPS satellites by timing a signal's journey from satellite to receiver. As it turns out, this is a fairly elaborate process.

At a particular time (let's say midnight), the satellite begins transmitting a long, digital pattern called a pseudo-random code. The receiver begins running the same digital pattern also exactly at midnight. When the satellite's signal reaches the receiver, its transmission of the pattern will lag a bit behind the receiver's playing of the pattern.

The length of the delay is equal to the signal's travel time. The receiver multiplies this time by the speed of light to determine how far the signal traveled. Assuming the signal traveled in a straight line, this is the distance from receiver to satellite.

In order to make this measurement, the receiver and satellite both need clocks that can be synchronized down to the nanosecond. To make a satellite positioning system using only synchronized clocks, you would need to have atomic clocks not only on all the satellites, but also in the receiver itself. But atomic clocks cost somewhere between $50,000 and $100,000, which makes them a just a bit too expensive for everyday consumer use.

The Global Positioning System has a clever, effective solution to this problem. Every satellite contains an expensive atomic clock, but the receiver itself uses an ordinary quartz clock, which it constantly resets.

In a nutshell, the receiver looks at incoming signals from four or more satellites and gauges its own inaccuracy. In other words, there is only one value for the "current time" that the receiver can use. The correct time value will cause all of the signals that the receiver is receiving to align at a single point in space. That time value is the time value held by the atomic clocks in all of the satellites. So the receiver sets its clock to that time value, and it then has the same time value that all the atomic clocks in all of the satellites have. The GPS receiver gets atomic clock accuracy "for free."

When you measure the distance to four located satellites, you can draw four spheres that all intersect at one point. Three spheres will intersect even if your numbers are way off, but four spheres will not intersect at one point if you've measured incorrectly. Since the receiver makes all its distance measurements using its own built-in clock, the distances will all be proportionally incorrect.

The receiver can easily calculate the necessary adjustment that will cause the four spheres to intersect at one point. Based on this, it resets its clock to be in sync with the satellite's atomic clock. The receiver does this constantly whenever it's on, which means it is nearly as accurate as the expensive atomic clocks in the satellites.

In order for the distance information to be of any use, the receiver also has to know where the satellites actually are. This isn't particularly difficult because the satellites travel in very high and predictable orbits. The GPS receiver simply stores an almanac that tells it where every satellite should be at any given time. Things like the pull of the moon and the sun do change the satellites' orbits very slightly, but the Department of Defense constantly monitors their exact positions and transmits any adjustments to all GPS receivers as part of the satellites' signals.

In the next section, we'll look at errors that may occur and see how the GPS receiver corrects them.

Differential GPS

So far, we've learned how a GPS receiver calculates its position on earth based on the information it receives from four located satellites. This system works pretty well, but inaccuracies do pop up. For one thing, this method assumes the radio signals will make their way through the atmosphere at a consistent speed (the speed of light). In fact, the Earth's atmosphere slows the electromagnetic energy down somewhat, particularly as it goes through the ionosphere and troposphere. The delay varies depending on where you are on Earth, which means it's difficult to accurately factor this into the distance calculations. Problems can also occur when radio signals bounce off large objects, such as skyscrapers, giving a receiver the impression that a satellite is farther away than it actually is. On top of all that, satellites sometimes just send out bad almanac data, misreporting their own position.

Differential GPS (DGPS) helps correct these errors. The basic idea is to gauge GPS inaccuracy at a stationary receiver station with a known location. Since the DGPS hardware at the station already knows its own position, it can easily calculate its receiver's inaccuracy. The station then broadcasts a radio signal to all DGPS-equipped receivers in the area, providing signal correction information for that area. In general, access to this correction information makes DGPS receivers much more accurate than ordinary receivers.

The most essential function of a GPS receiver is to pick up the transmissions of at least four satellites and combine the information in those transmissions with information in an electronic almanac, all in order to figure out the receiver's position on Earth.

Once the receiver makes this calculation, it can tell you the latitude, longitude and altitude (or some similar measurement) of its current position. To make the navigation more user-friendly, most receivers plug this raw data into map files stored in memory.

You can use maps stored in the receiver's memory, connect the receiver to a computer that can hold more detailed maps in its memory, or simply buy a detailed map of your area and find your way using the receiver's latitude and longitude readouts. Some receivers let you download detailed maps into memory or supply detailed maps with plug-in map cartridges.

A standard GPS receiver will not only place you on a map at any particular location, but will also trace your path across a map as you move. If you leave your receiver on, it can stay in constant communication with GPS satellites to see how your location is changing. With this information and its built-in clock, the receiver can give you several pieces of valuable information:

- How far you've traveled (odometer)
- How long you've been traveling
- Your current speed (speedometer)
- Your average speed
- A "bread crumb" trail showing you exactly where you have traveled on the map
- The estimated time of arrival at your destination if you maintain your current speed

Reference:

Brain, Marshall, and Tom Harris. "How GPS Receivers Work" 25 September 2006. HowStuffWorks.com. <http://electronics.howstuffworks.com/gadgets/travel/gps.htm> 24 August 2013.